Published in 49th Air Traffic Control Association Annual Conference Proceedings 2004

Using XML as a Common Format for the Exchange of Scenario Data

Robert D. Oaks
General Dynamics/Signal Solutions
robert.ctr.oaks@faa.gov

Mike Paglione
Federal Aviation Administration
mike.paglione@faa.gov

Robert D. Oaks is a Senior Computer Systems Analyst and Mike Paglione is the Simulation and Analysis Group's Conflict Probe Assessment Team Lead. Both are located at the FAA's William J. Hughes Technical Center, Atlantic City, New Jersey, 08405.

Introduction

Traffic flow management Decision Support Tools (DST), developed for the Federal Aviation Administration (FAA), use simulation as a tool for development, technical assessment, and field evaluation. The air traffic scenarios, used by these simulations, are files containing data describing the flow of aircraft traffic through an airspace over a period of time. The files contain both planning/advisory information and processed surveillance radar data, referred to as track. The time-stamped planning/advisory data describe each aircraft's planned flight. This includes each aircraft's flight plan and any flight plan amendments, interim altitude clearances, and hold information. The track data represents each aircraft's actual flight path. It consists of a series of 4-dimensional components: two defining the aircraft's position (e.g. latitude and longitude); another component defining its altitude; and the fourth component defining an associated time.

The Conflict Probe Assessment Team (CPAT) within the FAA's Simulation and Analysis Group has been evaluating FAA DSTs for aircraft trajectory and conflict prediction accuracy since 1996. In particular, a class of a DST, called a conflict probe, generates future predictions of conflicts (i.e. violations of standard separations) between aircraft and aircraft and restricted airspace. To evaluate the accuracy of these predictions, it is necessary to quantify the number of conflicts that are not predicted, referred to as missed alerts, and the number of non-conflicts that were predicted, referred to as false or nuisance alerts. To perform such an evaluation requires air traffic scenarios based on recorded field data that contain a significant number of conflicts. Since the scenario is based on recorded data, they contain the characteristics found in the operational environment. However, scenarios generated directly from the recorded field data do not contain conflicts, since air traffic controllers managed the airspace to avoid such events. Therefore, it was necessary for CPAT to modify the data to induce a number of conflicts, while at the same time retaining the realism of the original air traffic situation.

CPAT Scenario Generation Process

CPAT developed a scenario generation process consisting of three steps: data extraction, data modification, and scenario generation. During the data extraction step, tables in a relational database system are populated with data extracted from recorded traffic messages. During the data modification step the data in these tables may be modified.

During the scenario generation step, an air traffic scenario file is created using the table data. This CPAT scenario generation process has been documented in [Oaks and Paglione, 2001], [Oaks and Paglione, 2002], and [Paglione et al., 2003].

The CPAT scenario generation process evolved to satisfy changing needs. Its original purpose was to develop scenarios to support the accuracy testing of the User Request Evaluation Tool (URET), which is Lockheed Martin's implementation of the conflict probe originally developed by MITRE's Center for Advanced Aviation System Development (CAASD). The input files used for this original effort were data collected at URET Daily Use sites using a format called SCN established by MITRE CAASD. This is a proprietary format based on the FAA's 3.20 Host Computer System (HCS) Patch. The scenario files generated by the CPAT scenario generation process at this time used both this SCN format and a CMS_ASCII format, a proprietary format co-developed by CPAT and Lockheed Martin. CMS refers to Common Message Set, which is the FAA format used for interprocess communications between the HCS and collocated DSTs. As URET was deployed in the field, recorded data became available using this CMS_ASCII format, so the CPAT scenario generation process was modified to accommodate this format as well. While continuing to support the URET deployment, CPAT was also working closely with the NASA Ames Research Center and projects were proposed that would use yet another format called CM_SIM, which is NASA's proprietary format used to evaluate their Center-TRACON Automation System. More recently, the CPAT scenario generation process is expected to support testing for the En Route Automation Modernization project, which may require additional scenario formats to be adapted.

As the CPAT scenario generation process evolved to support these many testing efforts, it became obvious that the process needed to be rewritten to be more flexible. It also became obvious that a standard format for the exchange of scenario data between processes needed to be developed in order to keep up with the ever-changing systems with which CPAT had become involved. The high level design of this upgrade is depicted in Figure 1.

The processes shown on the left side of Figure 1 (*LMcmsToXml*, *ScnToXml*, and *CmSimToXml*) represent software involved with the data extraction step. These processes extract data from various data recording sources and produce a file identified as the *CMS XML file*. The processes on the right side of Figure 1 (*XmlToLMcms*, *XmlToScn*, and *XmlToCmSim*) represent software involved with the scenario generation step. These processes generate scenarios in formats appropriate for testing various systems. The process identified as *XmlToDb* is also a part of the data extraction step and populates the Air Traffic Database based on the contents of the *CMS XML file*. The process identified as *DbToXml* is a part of the scenario generation step and is used to create a *CMS XML file* from the contents of the tables in the Air Traffic Database. Not shown in Figure 1 are any of the processes involved with the data modification step. These processes interface with the tables in the Air Traffic Database and are not discussed in this paper. An example of scenario modification is presented in [Paglione et al., 2003]. The format of the *CMS XML file* is discussed in this paper, since it represents the common exchange format or "core currency" selected by CPAT for their upgraded scenario generation process.

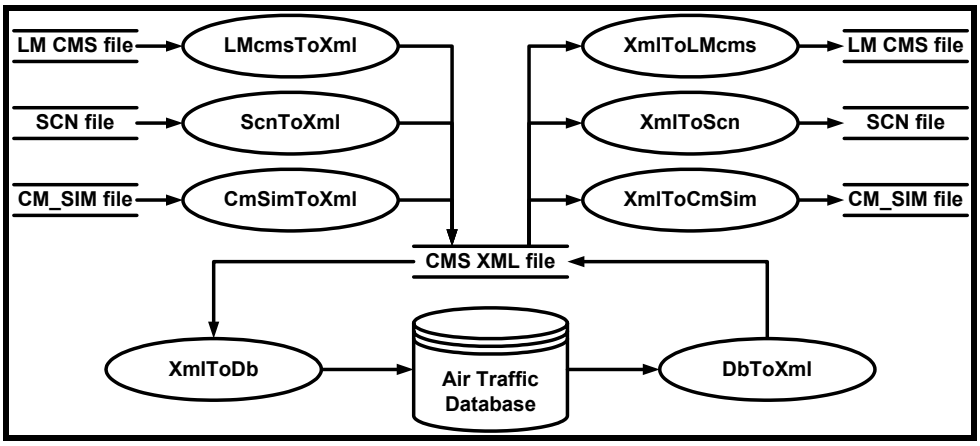

Figure 1: High Level Design of the Upgraded CPAT Scenario Generation Process

Extensible Markup Language

The *CMS XML file* uses a format based on the Extensible Markup Language (XML). This is a metamarkup language in which data is presented as strings of text surrounded by text markup that describes the enclosed data. XML is called a metamarkup language because it does not have a fixed set of tags and elements that are designed to operate for any application. XML is called extensible because the language can be extended and adapted to meet many different needs. As a result, XML allows its users to define the tags and elements they need as they need them. Because of this flexibility, it is playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. The technical specifications for XML are published by the World Wide Web Consortium (W3C). These specifications and other information about XML can be found on their web site (www.w3c.org). A large number of books are available that describe XML and its applications. One especially good resource, which is used by CPAT, is *The XML CD Bookshelf* published by O'Reilly & Associates [O'Reilly, 2002]. CPAT selected XML as the common exchange format for a number of reasons including:

- First, XML is portable. An XML-formatted data file can be moved easily between various platforms because it is a simple ASCII text format. This means that the scenarios generated by the CPAT scenario process can be used on a Solaris system using a UNIX operating system or on a personal computer using either a Windows or LINUX operating system.

- Next, XML is a well-defined standard. An XML-formatted data file must conform to the specification defined by the W3C. Therefore, it is easy to specify the requirements for the software that processes the scenario files.

- In addition, a large amount of support software exists. In particular, this includes a well-established Java API. This is especially important since CPAT also decided to rewrite its scenario generation process using the Java programming language.

- Furthermore, the data allows for the maximum error checking, since the well-defined XML standard includes capabilities to specify the syntax of the scenario messages. Both CPAT's parsers and commercially off-the-shelf software tools can verify that all messages comply with the specified syntax.

- Finally, the data for our specific purpose is easily represented using XML, especially now with the established Common Message Set (CMS) as the standard format for messages exchanged with Air Traffic Management Tools.

Common Message Set

The XML format implemented by CPAT is based strictly on the CMS message formats specified in the CMS Interface Requirements Document (IRD) [FAA, 2000]. This document defines the data exchanged between the HCS and various DSTs. An example of a scenario file using this format is presented in Figure 2. In this example it is seen that the tags for the interprocess messages are the same as the message identifier defined in the CMS IRD; for example, FH for Flight Plan Information message and TH for Track Information message. It is also seen that the tags for the data contained in the messages are based on the Field Reference Number defined in the CMS IRD; for example, REF008A refers to Field Reference Number 08a, which is defined in Table 30-I of the CMS IRD. In this example the <REF008A>350</REF008A> refers to Field Reference Number 08, which is the Assigned Altitude specified using Field Format Id "a", which is 2 to 3 digits representing altitude or flight level. CPAT chose this convention purposefully so the tags used in this XML scenario format are precisely defined by an FAA standard, the CMS IRD. This does not create a readability problem, since an XML file is meant for the exchange of data between software processes rather than a direct presentation to a human.

References

[FAA, 2000] Federal Aviation Administration (2000), "Interface Requirements Document ARTCC Host Computer System (HCS) / Air Traffic Management (ATM) Applications," NAS-IR-82170001, September 1, 2000, Revision M.

[Oaks and Paglione, 2001] Oaks, R. D. and M. M. Paglione, "Generation of Realistic Air Traffic Scenarios Based on Recorded Field Data," *46th Annual Meeting of the Air Traffic Control Association*, Washington, DC, November, 2001.

[Oaks and Paglione, 2002] Oaks, R. D. and M. M. Paglione, "Generation of Realistic Air Traffic Scenarios Using a Genetic Algorithm," 21st Digital Avionics Systems Conference, Irvine, CA, October, 2002.

[O'Reilly, 2002] O'Reilly & Associates, *The XML CD Bookshelf, Version 1.0*, ISBN: 0-596-00335-8, O'Reilly & Associates, Inc., Sebastopol, CA 95472.

[Paglione et al., 2003] Paglione, Mike M., Robert D. Oaks, and Karl D. Bilimoria (2003), "Methodology for Generating Conflict Scenarios by Time Shifting Recorded Traffic Data," *AIAA 3rd Annual Aviation Technology, Integration, and Operations Technical Forum*, Denver, CO.

```
<?xml version="1.0">
<CPAT_CMS_FILE>
 <DATE>20030611</DATE>
 <START_TIME>150000973</START_TIME>
 <FH>
  <TIMESTAMP>150000973</TIMESTAMP>
  <SEQ>1</SEQ>
  <SRC_ADDR>ZZZHOST0</SRC_ADDR>
  <DEST_ADDR>URETZZZ0</DEST_ADDR>
  <REF002A>ABC1234</REF002A>
  <REF002D>321</REF002D>
  <REF003A>1</REF003A>
  <REF003C>MD80</REF003C>
  <REF003E>A</REF003E>
  <REF004A>1354</REF004A>
  <REF005A>449</REF005A>
  <REF006A>MSL213038</REF006A>
  <REF007D>E1500</REF007D>
  <REF008A>350</REF008A>
  <REF010A>RSW./.MSL213038..SGF.TRAKE8.STL/1544</REF010A>
  <REF143A>3201</REF143A>
  <REF149A>EOM</REF149A>
 </FH>
...
 <TH>
  <TIMESTAMP>150007975</TIMESTAMP>
  <SEQ>12</SEQ>
  <SRC_ADDR>ZZZHOST0</SRC_ADDR>
  <DEST_ADDR>URETZZZ0</DEST_ADDR>
  <REF002A>ABC1234</REF002A>
  <REF002D>321</REF002D>
  <REF005B>436</REF005B>
  <REF008A>350</REF008A>
  <REF054A>350</REF054A>
  <REF138A>ZCM</REF138A>
  <REF138B>32</REF138B>
  <REF023D>341040N/0875500W</REF023D>
  <REF023E>-371/+224</REF023E>
  <REF149A>EOM</REF149A>
 </TH>
...
 <END_TIME>160001376</END_TIME>
</CPAT_CMS_FILE>
```

Figure 2: Sample XML Scenario File